AD-A271 520 ON PAGE

1. A			3. REPORT TYPE AND DATES COVERED
	August 1993		Final 15 Mar 90-30 Sep 93

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Functional Statistical Data Analysis and Modelling	DAAL03-90-G-0069

6. AUTHOR(S)

Emanuel Parzen

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Department of Statistics Texas A&M University College Station, TX 77843-3143	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
U.S. Army Research Office P. O. Box 12211 Research Triangle Park, NC 27709-2211	ARO 27574.16-MA

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Approved for public release; distribution unlimited.	

13. ABSTRACT *(Maximum 200 words)*

Our research during 1990-1993 was reported in 13 published papers and one Ph.D.
thesis listed below. We regard our research as concerned with (numbers in
parentheses represent number of papers in the area):

1. Change Analysis (4)
2. Functional Inference (4)
3. Time Series Analysis (2)
4. Statistical Culture and History (3)
5. Supervision of Ph.D. theses

93 10 15 220

93-24674

14. SUBJECT TERMS			15. NUMBER OF PAGES
			9
			16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

SUMMARY OF WORK ACCOMPLISHED

During the years 1990–1993 our research project "Functional Statistical Data Analysis and Modeling" was unusually productive because our research on Functional Statistical Methods (based on the unifying idea of comparison density functions) culminated in, and was clarified by, the development of a new research area which we call Change Analysis; it is an extension of Change-Point estimation which has an extensive literature.

Change detection and estimation should emerge as a major research field at the leading edge of research in theoretical and applied statistics; evidence of this is the several major workshops in 1992-1993 (AMS-IMS-SIAM, Carleton University, University of Maryland) in which we actively participated, and in whose proceedings we have papers.

Our research during 1990–1993 was reported in 13 published papers and one Ph.D. thesis listed below. We regard our research as concerned with (numbers in parentheses represent number of papers in the area):

1. Change Analysis (4)

2. Functional Inference (4)

3. Time Series Analysis (2)

4. Statistical Culture and History (3)

5. Supervision of Ph.D. theses.

Change analysis, and change-point estimation, extend standard statistical methods (which estimate the parameters of a probability model for a data set) by testing if there is a change in the parameters over the data, and modeling how probability distributions are changing. Our approach to change analysis extends tests for independence of two random variables X and Y; in change problems Y is the observation and X is the index of observation. Our papers introduce the theory and practice of a new non-parametric approach called the Comparison Change approach, and a new parametric approach called Fisher Score Change Processes. An appendix outlines the new functions introduced in our theory, and our current research on applying change concepts to the presentation of

2

standard statistical methods and the development of "beyond AOV" techniques.

Functional inference adds to standard statistical methods data analysis methods based on quantile domain functions, information measures, and comparison density functions. Important results in our papers include: new goodness of fit tests based on entropy and comparison density functions; unification of statistical methods for continuous and discrete data enhanced by graphical methods inspired by change analysis; Change PP plot and continuous versions of sample quanntile functions.

Time series analysis is still an extremely active and important interdisciplinary research area. Recent developments are described in two recent major collections of papers which include papers by us. We continued to provide leadership in this field by organizing the U.S. Japan Joint Seminar on Statistical Time Series Analysis, Honolulu, Hawaii. January 24–29, 1993.

Statistical culture and history is important (it should not be neglected by statisticians in an era where change, and continuous reviews of research directions to determine needed changes or reforms, is increasingly normal). Our papers discuss: introductory probability in the era of the 1960's as reflected in textbooks by Mosteller; the career of Hirotugu Akaike (a role model for the conduct of interdiscplinary statistical research); the concept of "hammers and nails" in the practice of statistics; proposals for improving the impact of statistics.

In the area of supervision of Ph.D. theses, Cheng Cheng received his Ph.D. in August 1993 with research in the areas of our research program (he is currently with Upjohn Laboratories in Kalamazoo, Michigan, as a post-doctoral fellow in molecular similarity). Cheng's Ph.D. thesis, entitled *On estimation of Quantile and Quantile Density Functions*, has 300+ pages and presents major results in Functional Statistical Inference. It is outstanding in scholarship (it provides a comprehensive review of many quantile estimators proposed in the literature) and original research (it proposes new estimators for practical estimation). In the pipe-line are two Ph. D. students (Todd Ogden and Ying-Sheng

3

Hu) who are currently doing reseasrch on change analysis estimation methods which use function approximation by *wavelets*.

APPENDIX

A. Comparison Change Analysis Basic Identities Formal Derivation

We believe that the relations between various comparison functions can best be understood by formally deriving them from the conditional distribution identity

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} F_{Y|X=x'}(y)dF_X(x').$$

The independence testing function $D(\tau, u) = F_{X,Y}(Q_X(\tau), Q_Y(u))$ has representation

$$F_{X,Y}(Q_X(\tau), Q_Y(u)) = \int_{-\infty}^{Q_X(\tau)} F_{Y|X=x'}(Q_Y(u))dF_X(x')$$

Make change of variable $t = F_X(x')$, $x' = Q_X(t)$. Assume τ, u exact. Then

$$F_{X,Y}(Q_X(\tau), Q_Y(u)) = \int_0^{\tau} F_{Y|X=Q_X(t)}(Q_Y(u))dt.$$

A similar derivation shows

$$F_{X,Y}(Q_X(\tau), Q_Y(u)) = \int_0^{u} F_{X|Y=Q_Y(u')}(Q_X(\tau))du'.$$

Comparison functions

$$D(t; u) = F_{Y|X=Q_X(t)}(Q_Y(u))$$

$$D([0, \tau]; u) = F_{Y|X \leq Q_X(\tau)}(Q_Y(u))$$

satisfy Change Analysis Identities:

$$D(\tau, u) = \tau D([0, \tau]; u)$$

$$= \tau \int_0^{u} d([0, \tau]; u')du'$$

$$D(\tau, u) = \int_0^{\tau} D(t; u)dt$$

$$= \int_0^{\tau} dt \int_0^{u} du' d(t, u')$$

4

assuming in discrete case that τ and u are "exact" values.

Comparison densities satisfy

$$d(t, u) = d(u; F_Y, F_{Y|X=Q_X(t)})$$

$$= d(t; F_X, F_{X|Y=Q_Y(u)})$$

$$\tau d([0, \tau], u) = \int_0^\tau d(t, u)dt$$

$$\tau d([0, \tau], u) = P[X \le Q_X(\tau)|Y = Q_Y(u)]$$

$$= F_{X|Y=Q_Y(u)}(Q_X(\tau))$$

B. COMPARISON CHANGE AND QUANTILE ANALYSIS APPROACH TO UNIFICATION OF STATISTICAL METHODS AND BEYOND AOV TECHNIQUES

Introduction

This paper describes the new concepts of our research (Parzen (1979), (1989), (1992), (1993)) on a comprehensive development of change and quantile analysis function smoothing methods for statistical data analysis and tests of homegeneity of univariate, bivariate, and multi-sample data sets. The aim of the paper is to outline methods and techniques we have developed, summarize proposed graphs, discuss examples, and hopefully demonstrate that a theory extending statistical methods provides benefits of (1) Utility: providing new methods which may be break throughs, in a sense that they help solve problems untouched by standard methods, and (2) Beauty: enhance understanding of standard methods.

Our research goal is to unify conventional and new statistical methodologies, such as those which Hirotsu (1993) calls "beyond analysis of variance (AOV) techniques". We propose a framework which integrates conventional (AOV and non-parametric) and emerging (beyond AOV) statistical techniques.

The paper has several parts: I. Univariate quantile data analysis; II. Univariate Change-point analysis; III. Multisample data analysis. EXPLORE: an S-plus and Fortran library of computer programs for data analysis developed by Cheng Cheng is described in an appendix. Figures of functional statistics are all at the end of the paper.

5

PUBLICATIONS OF
ARO CONTRACT DAAL03-90-G-0069

Change Analysis

Emanuel Parzen. (1992). 'Comparison Change Analysis'. *Nonparametric Statistics and Related Topics* (ed. A. K. Saleh), Elsevier: Amsterdam, 3-15.

Emanuel Parzen. (1993). 'Change Analysis' *Proceedings of the 36th Army Conference on the Design of Experiments*, 225-232.

Emanuel Parzen. (1993). 'Limit Theorems for Fisher-Score Change Processes,' *Changepoint Analysis Lecture Notes in Statistics*, Institute of Mathematical Statistics. (Joint with Lajos Horváth).

Emanuel Parzen. (1993). 'Comparison Change Analysis Approach to Changepoint Estimation,' *Applied Changepoint Analysis Symposium, Journal of Applied Stati.. 'ral Science.*

Emanuel Parzen. (1993). 'Comparison Change Analysis and Empirical Processes', manuscript.

Emanuel Parzen. (1993). 'Comparison Change and Quantile Analysis Approach to Unification of Statistical Methods and Beyond AOV Techniques', manuscript.

Functional Inference

Emanuel Parzen. (1991). 'Goodness of Fit Tests and Entropy,' *Journal of Combinatorics, Information, and System Science*, 16, 129-136.

Emanuel Parzen. (1991). 'Unification of Statistical Methods for Continuous and Discrete Data,' *Proceedings Computer Science-Statistics INTERFACE '90*, (ed. C. Page and R. LePage), Springer Verlag: New York, 235-242.

Emanuel Parzen. (1993). 'From Comparison Density to Two Sample Data Analysis,'

The Frontiers of Statistical Modeling: An Informational Approach, ed. H. Bozdogan, Kluwers: Amsterdam.

Emanuel Parzen. (1993). 'Change *PP* Plot and Continuous Sample Quantile Function,' *Communications in Statistics*.

Time Series Analysis

Emanuel Parzen. (1992). 'Time Series, Statistics, and Information,' *New Directions in Time Series Analysis* (ed. E. Parzen et al). Springer Verlag: New York, 265-286.

Emanuel Parzen, (1993) 'Stationary Time Series Analysis Using Information and Spectral Analysis,' *Developments in Time Series Analysis (in Honor of Maurice Priestley)*. Ed. T. Subba Rao. Chapman Hall, 139-148.

Statistical Culture and History

Emanuel Parzen. (1991). 'Probability with Statistical Applications,' a contribution to *A Statistical Model: Frederick Mosteller*, edited by S. E. Fienberg and D. C. Hoaglin. Springer Verlag: New York, 199-200.

Emanuel Parzen. (1993). 'History of Statistics in Real Time: Hammers and Nails', *Computer Science-Statistics, Interface Foundation*, , Vol. 24, 602-608.

Emanuel Parzen. (1993). 'Hirotugu Akaike, Statistical Scientist,' *TheFrontiers of Statistical Modeling: An Informational Approach*, ed. H. Bozdogan, Kluwers: Amsterdam.

Ph.D. Thesis

Cheng Cheng. (1993). 'On Estimation of Quantile and Quantile Density Functions.' Also, Technical Report #200 and #201, Department of Statistics, Texas A&M University.

ABSTRACT

On Estimation of Quantile and Quantile Density Functions. (August 1993)

Cheng Cheng, B.S., Beijing Computer Institute;

M.S., University of Texas at El Paso

Chair of Advisory Committee: Dr. Emanuel Parzen

In statistical analysis the quantile function (qf) and the quantile density function (qdf) are equally important as the cdf and the pdf. Numerous smooth quantile function estimators have been proposed as alternatives to the sample quantile function. The research in this area has been somewhat esoteric in the sense that little work has been done in terms of providing guidelines for practice. This research is motivated by several practical issues involved in quantile function estimation.

In the attempt to compare various qf estimators and to provide guidelines for their uses in practice, asymptotic behavior of translation and scale equivariant qf estimators and the derived qdf estimators are first investigated under a unified representation. It is shown that under moderate conditions these estimators share equally good asymptotic behavior. Some existing asymptotic results are strengthened and extended. Sufficient conditions for the consistency of the derived qdf estimators are obtained as well.

The practical impact of the asymptotic results is the demonstration of the fact that smoothing the sample quantile function by any reasonable kernel always results in a qf estimator possessing nice asymptotic behavior. So asymptotic theory is not informative for obtaining guidelines for the selection of kernels in practice (finite samples). Finite-sample behavior of several translation and scale equivariant qf estimators is then formulated and compared according to several criteria. The finite sample results, in conjunction with the asymptotics, clearly suggest certain smoothing kernels to use in practice.

A data-driven procedure to determine amount of smoothing is proposed. State-of-the-art demonstrations of the quantile-domain data analysis methodology are presented by analyzing several well-known data sets.